

Georeferencing Textual Annotations and Tagsets with Geographical Knowledge and Language Models

Georeferenciado de Anotaciones Textuales y Etiquetas Semánticas con Conocimiento Geográfico y Modelos de Lenguaje.

Daniel Ferrés

TALP Research Center
Universitat Politècnica de Catalunya
C/ Jordi Girona Salgado 1-3
08034 Barcelona
dferres@lsi.upc.edu

Horacio Rodríguez

TALP Research Center
Universitat Politècnica de Catalunya
C/ Jordi Girona Salgado 1-3
08034 Barcelona
horacio@lsi.upc.edu

Resumen: Presentamos en este artículo cuatro aproximaciones al georeferenciado genérico de anotaciones textuales multilingües y etiquetas semánticas. Las cuatro aproximaciones se basan en el uso de 1) Conocimiento geográfico, 2) Modelos del lenguaje (LM), 3) Modelos del lenguaje con predicciones re-ranking y 4) Fusión de las predicciones basadas en conocimiento geográfico con otras aproximaciones. Los recursos empleados incluyen el gazetteer geográfico Geonames, los modelos de recuperación de información TFIDF y BM25, el Hiemstra Language Modelling (HLM), listas de stop words para varias lenguas y un diccionario electrónico de la lengua inglesa. Los mejores resultados en precisión del georeferenciado se han obtenido con la aproximación de re-ranking que usa el HLM y con su fusión con conocimiento geográfico. Estas estrategias mejoran los mejores resultados de los mejores sistemas participantes en la tarea oficial de georeferenciado en MediaEval 2010. Nuestro mejor resultado obtiene una precisión de 68.53% en la tarea de georeferenciado hasta 100 Km.

Palabras clave: Georeferenciado, Gazetteers Geográficos, Recuperación de la Información, Modelos de Lenguaje.

Abstract: This paper describes generic approaches for georeferencing multilingual textual annotations and sets of tags from metadata associated to textual or multimedia content with high precision. We present four approaches based on: 1) Geographical Knowledge, 2) Language Modelling (LM), 3) Language Modelling with Re-Ranking predictions, 4) Fusion of Geographical Knowledge predictions with the other approaches. The resources employed were the Geonames geographical gazetteer, the TFIDF and BM25 Information Retrieval algorithms, the Hiemstra Language Modelling (HLM) algorithm, stopwords lists from several languages, and an electronic English dictionary. The best results in georeferencing accuracy are achieved with the HLM Re-Ranking approach and its fusion with Geographical Knowledge. These strategies outperformed the best results in accuracy reported by the state-of-the art systems that participated at MediaEval 2010 official Placing task. Our best results achieved are 68.53% of accuracy georeferencing up to a distance of 100 Km.

Keywords: Georeferencing, Geographical Gazetteers, Information Retrieval, Language Models.

1 Introduction

Due to the increasing amount of annotated content in social platforms on the web like blogs (blogspot, wordpress, ...), social networks (Facebook, Twitter,...), and multimedia sharing platforms (Flickr, Panoramio, Youtube,...) there is a growing need of ap-

plications that can deal with automatic understanding of users textual annotations and tags. The approach presented in this paper deals with automatic understanding of geographical place names in users textual annotations (including tags) and georeferencing this information. Geotagging (Kefler et al., 2009) or georeferencing consists of ex-

tending the information of tags by predicting an explicit location in space and time¹ to a textual or multimedia content. The Georeferenced textual or multimedia content contains the geographical coordinates of the location where the content refers. Currently some platforms allow users to georeference (geotag) their content automatically (GPS-enabled cameras) or manually. Unfortunately most of the textual and media content is not georeferenced, and this implies the need of automatic georeferencing.

In this paper we present our approaches, experiments, and conclusions in georeferencing users textual annotations and tagsets. The experiments were performed with the MediaEval 2010 Placing task² data sets. The MediaEval 2010 Placing task (Larson et al., 2011) required that participants automatically assign geographical coordinates (latitude and longitude) to Flickr videos using one or more of: Flickr metadata, visual content, audio content, and social information³. This data set gives the researchers the opportunity to deal with multilingual non grammatical language issues appearing in the users textual annotations and geographical disambiguation and focusing strategies.

Our Geographical Knowledge based approach to the textual georeferencing task is to deal with automatic understanding of geographical textual annotations and geographical tags by applying the tasks of Place Name Recognition (detects the place names in the textual annotations and tagsets) and Geographical Focus Detection (decides which place name is the focus of the content and retrieves its coordinates). This approach has the objective of maximising the precision of the predicted places at the expense of recall, predicting only videos with textual annotations that contain geographical place names detected in the Place Name Recognition phase. Our experiments with the Geographical approach include the use stopwords lists and controlled dictionaries for the disambiguation of placing names and the focus detection. We decided to explore approaches with Language Models with and without Re-Ranking that have been used

successfully in georeferencing ((Serdyukov, Murdock, and van Zwol, 2009), (Hays and Efros, 2008), (Laere, Schockaert, and Dhoedt, 2010b)). Finally we explored the GeoFusion approach that combines Geographical Knowledge and LM in order to test the combination of both approaches.

This paper is structured as follows. In the next section we review the related research in georeferencing. In Section 3 we describe our geographical knowledge based approach to the textual annotations and tags georeferencing. In Section 4 we expose our georeferencing experiments with the geographical approach and their results with the MediaEval 2010 Placing task test set. In Section 5 we describe our approach based on Language Models and present some experiments with them. In Section 6 we introduce an approach based on Re-Ranking the results obtained by the Language Models, and another approach that combines the Geographical approach with the other approaches, we also present there the experiments with these approaches. Finally, in Section 7 we discuss and report our conclusions in the context of georeferencing, and we develop new ideas for further research for improving georeferencing tasks.

2 Related Work

Many approaches to the Georeferencing task for predicting the most appropriate coordinates for tagged images have been presented in last years using mainly visual and/or textual features. Also some of the georeferencing approaches use its own and different corpus collections, and another ones within the MediaEval 2010 evaluation. (Crandall et al., 2009) presented a system that uses textual (tags), visual, and temporal features for placing Flickr images on map. They used automatic classifiers based on Naïve Bayes and Support Vector Machines trained over a corpus of 35 million images. Their results show that visual and textual features together outperform text features alone by a significant margin. (Serdyukov, Murdock, and van Zwol, 2009) used a language model based on the tags provided by the users to predict the location of Flickr images. The language model follows a Multinomial distribution. In addition several smoothing strategies were taken into account to test: 1) spatial neighbourhood for tags, 2) cell relevance

¹In this paper we only consider georeferencing in space.

²www.multimediaeval.org

³The approaches presented in this paper use only metadata information.

probabilities, 3) toponym-based smoothing, 4) spatial ambiguity-aware smoothing. They used a set of 140,000 geo-tagged Flickr photos in which there is at most one photo per user with the same tagset. They used 120,000 to train models, 10,000 for tuning parameters and 10,000 for testing purposes. All the smoothing strategies outperformed the Language Modelling baseline. (Laere, Schockaert, and Dhoedt, 2010b) presented a system that uses clustering (k-medoids) and classification (Naïve Bayes) algorithms to predict geographic coordinates of Flickr photos by using users tags. (Hays and Efros, 2008) used only visual features and dataset of over 6 million GPS-tagged images from the Flickr online photo collection to estimate the geographical coordinates of the image. Their results were evaluated over a set of 237 images. They use a measure of visual features similarity between test image and dataset image to find the 1-NN (Nearests Neighbours) and the 120-NN. Mean-shift clustering is then applied to the estimated probability distribution derived from the 120-NN photos with GPS-coordinates. They report results in prediction locations up to 30 times better than chance.

Our approach is completely text-based and adapted to the analysis of tags and this means that like other approaches based on Language Modelling (Serdyukov, Murdock, and van Zwol, 2009), (Laere, Schockaert, and Dhoedt, 2010b) can be easily adapted to different types of textual annotations and sets of tags associated to content (textual or multimedia) from different social networks.

2.1 Georeferencing at MediaEval 2010

Systems presented at Mediaeval2010 such as (Laere, Schockaert, and Dhoedt, 2010a), (Kelm, Schmiedeke, and Sikora, 2010), (Ferrés and Rodríguez, 2010), and (Perea-Ortega et al., 2010), offered a possibility to compare their results over the same corpus. (Laere, Schockaert, and Dhoedt, 2010a) obtained the best results at the MediaEval 2010 placing task obtaining a 67,23% of accuracy predicting georeferences up to 100 Km with a system that applies Language Modelling and Clustering. They used a corpus of 8,685,711 annotated metadata from Flickr photos to train the models. (Kelm, Schmiedeke, and Sikora,

2010) presented an approach at MediaEval 2010 that combines three different methods to estimate geographical regions: a natural language processing approach with geographical knowledge filtering, probabilistic latent semantic analysis (pLSA) document indexing, and classification method based on colour and edge visual features to train a support vector machine (SVM). In order to apply Natural Language Processing (NLP) processors they detected and translated the textual annotations to English using the Google translate service. Their NLP approach is based on the use of a NLP processor, Wikipedia⁴ and Geonames⁵ with the use of population and higher-level categories salience. They achieve their best results in accuracy for prediction at a maximum of 100 Km with the combination of NLP and Geographic Knowledge method and visual features with a result of 60.46% of accuracy. We (Ferrés and Rodríguez, 2010) presented a system that achieved an accuracy of 52% georeferencing up to a distance of 100 Km using only the Geonames Gazetteer and other lexicons. (Perea-Ortega et al., 2010) presented a system at MediaEval 2010 that uses an approach based on applying a geographical Named Entity Recognizer (Geo-NER) on the textual annotations. Geo-NER is a geographical entity recognizer that makes use of Wikipedia and GeoNames. A main difference from the text based approaches systems of (Kelm, Schmiedeke, and Sikora, 2010) and (Perea-Ortega et al., 2010) is that we do not use Named Entity Recognizers and NLP processors. Our system uses two geographical focus disambiguation strategies, one based on population heuristics and another that combines geographical knowledge and population heuristics.

3 Geographical Approach

The approach of using only Geographical Knowledge for georeferencing has two main phases: Place Names Recognition and Geographical Focus Disambiguation. Place Name Disambiguation is applied in both phases, with place name normalisation and geo/non-geo ambiguities solved in the first one and the geographical class ambiguities and reference ambiguity applied in the second one.

⁴**Wikipedia.** <http://www.wikipedia.org>

⁵**Geonames.** <http://www.geonames.org>

3.1 Place Names Recognition

The Place Names Recognition phase uses the Geonames Gazetteer for detecting the place names in the textual annotations. The Geonames Gazetteer contains over eight million geographical names and consists of 6.5 million unique features classified with a set 645 different geographical feature types. We decided not to use a NERC (Named Entity Recognition and Classifier) for several reasons: i) multilingual annotations complicate also the application of Natural Language tools such as POS-tagging, and NERC, ii) textual annotations and tags are not suitable for most NERC systems trained in news corpora. iii) some NERC systems are not performing much better than geographical Names Recognition from gazetteer lookup (Stokes et al., 2008). On the other hand we know also some limitations of a Gazetteer lookup approach: highly irregular coverage (Adrian Popescu, 2008), and poor spatial inclusion defined (Popescu and Kanellos, 2009).

The Geonames gazetteer has been used in Geographical Information Retrieval (GIR), Geographical Scope Resolution (Geoffrey Andogah, Gosse Bouma, John Nerbonne, Elwin Koster, 2008) and georeferencing approaches (Serdyukov, Murdock, and van Zwol, 2009), and allows us to deal with the recognition of issues related with multilinguality, acronyms, lower and uppercase place names, and word joined place names (e.g. riodejaneiro). We use the following information from each Geonames place: country, state, and continent of the place, feature type, coordinates, and population. The Place Names Disambiguation phase tries to avoid ambiguity in Place Names Recognition due to the huge number of place names with meaning that could be a noun (e.g. aurora (noun), Aurora (city)). This phase uses stopwords lists⁶ in several languages (including English) and an English Dictionary of 71.348 words get from Freeling⁷ toolbox to filter out names that could be erroneously tagged as place names.

3.2 Geographical Focus Detection

The Geographical Focus Detection phase uses some of the Toponym Resolution strate-

gies presented in the GIR literature (Leidner, 2007). This phase has been designed with some heuristics described in (Hauptmann, Hauptmann, and Olligschlaeger, 1999) and (Leidner, 2007). We assume the one reference per discourse hypothesis: one geographical place/coordinates per video and if there are no detected place names in the textual annotations the georeference is unresolved. Using the information of all possible referents of all the place names detected by the Geonames (phase 1) and not filtered (phase 2) we apply the following heuristics:

- **Population heuristics.** Only using population information to disambiguate between all the possible places. We used the following rules: 1) if a place exists select the most populated place that is not a country, state (administrative division type 1) or a continent, 2) otherwise if exists a state, select the most populated one, 3) otherwise select the most populated country, 4) otherwise select the most populated continent.
- **Geographical Knowledge heuristics.** These heuristics are similar to the toponym resolution algorithm applied by (Hauptmann, Hauptmann, and Olligschlaeger, 1999) to plot on a map locations mentioned in automatically transcribed news broadcasts. In our system the geo-class ambiguity between country names and city names (e.g. Brasil (city in Colombia) versus Brasil (country)) is resolved giving priority to the country names, and the geo-class ambiguity between state names and city names is resolved giving priority to city names. From the set of different places appearing in the text apply the following geographical scope detection and toponym disambiguation rules in order to select the scope (focus) of the text: H1) select the most populated place that is not a state, country or continent and has its state appearing in the text, H2) select the most populated place that is not a state, country or continent and has its country appearing in the text, H3) otherwise select the most populated state that has his country appearing in the text H4) otherwise do not resolve.
- **Geographical Knowledge and Population heuristics.** This method

⁶Lingua::StopWords 0.09.
<http://search.cpan.org/dist/Lingua-StopWords>

⁷Freeling 2.1. <http://nlp.lsi.upc.edu/freeling/>

combines the Geographical Knowledge and the Population heuristics presented above in the following way: 1) if it is possible apply the geographical knowledge heuristics, 2) otherwise apply the population heuristics presented above.

4 Georeferencing Experiments

The MediaEval 2010 Placing task data sets are composed by 5125 and 5091 videos (and its metadata) for the development and test sets respectively. All the experiments presented in this paper are realized over the MediaEval 2010 Placing task test set (consisting of 5091 metadata and videos). Evaluation of results is done by calculating the distance (haversine formula) from the actual point (assigned by a Flickr user) to the predicted point (assigned by a participant). Runs are evaluated finding how many videos were placed at least within some threshold distance: 1 km, 5km, 10km, 50km and 100km. We used only textual content from the following metadata fields from the Flickr videos to perform the task: Title, Description, and Keywords (set of tags). In our initial experiments at MediaEval 2010 (see (Ferrés and Rodríguez, 2010)) we allowed the Gazetteer the recognition of place names of a maximum five tokens (e.g. Sierra Nevada de Santa Marta), that were increased to seven in the new experiments presented in this paper (without improvements noticed).

The georeferencing experiments with the Geographical Knowledge approach presented in this paper are the following: 1) experiments to detect the relative importance of the metadata fields (title, description, and keywords) for the georeferencing task, 2) experiments to detect the performance and precision of the geographical disambiguation heuristics. The experiments performed with the geographical knowledge approach include the following improvements with respect our initial experiments with the MediaEval 2010 data set: 1) filter weak geographical named Entities (e.g. the geoname *Porto Alegre* (Brasil) has the weak geonames Porto and Alegre which could be matched as geonames but are not recognized), 2) improving toponym resolution prioritizing geographical salience of places with a country, state or continent appearing in the metadata over places with more population. As an example, for desambiguating the geoname “Paris”, Paris

(Texas) will be prioritized in the scope ranking algorithm over Paris (France) if “Texas” appears in the metadata), and 3) adding the Geonames Alternate Names file (with 2.9 million of features). The experiments show that the improvements of a weak NE filter and the geosalience can improve the results of the georeferencing. On the other hand the addition of the Alternate Names data does not improve the results. The improvement with respect of the best results obtained with the MediaEval 2010 data set is from 2740 videos (0.5382 of accuracy) to 2838 videos (0.5574 of accuracy) correctly georeferenced at a distance maximum of 100 Km.

Table 1: Experiments with different metadata fields.

Metadata fields	Accuracy (over 5091 videos)				
	1km	5km	10km	50km	100km
Title	0.034	0.076	0.090	0.104	0.109
Descript.	0.023	0.075	0.098	0.131	0.132
Keywords	0.106	0.331	0.412	0.503	0.520
T + K	0.103	0.336	0.421	0.517	0.537
T + D	0.051	0.135	0.168	0.210	0.217
K + D	0.108	0.343	0.429	0.526	0.544
T+K+D	0.105	0.345	0.433	0.532	0.552

See in Table 1 the results of the experiments to see the importance of the Metadata fields used in our approach: Title, Keywords, and Description (see Table 2 for an example). The results show that the metadata field Keywords (tags) is the most important one, achieving results of 52% of accuracy at 100 Km. The inclusion of the Title and/or the Description fields improves the results of the Keywords (tags) alone: Title and Keywords (53.7%), Keywords and Description (54.4%), Title and Keywords and Description achieves the 55.2% of accuracy at 100 Km. The use of only Title and Keywords achieves an accuracy of 21.7% at 100 Km this is slightly better than the results of the (Perea-Ortega et al., 2010), that used a Geo-NER for detecting Named Entities in title and Description achieving an accuracy of 21.3%. Although is not clear how to compare the toponym disambiguation process between the two systems this may indicate that Gazetteer lookup performance with Geonames is performing at state-of-the-art NERC level.

See in Table 3 the georeferencing experiments done to show the performance and importance of geographical knowledge and pop-

Table 2: Sample of Georeferencing metadata associated to a Flickr Video with the Geographical Knowledge approach.

ID VIDEO	00854
TITLE	Lensbaby D90 Video.
KEYWORDS	lonsdalequay, northvancouver, ocean, lensbaby, composer, d90, video, smartcookies
Prediction	49.31636 -123.06934 (northvancouver P PPL CA 02 48000 68 46)
Ground Truth	<Latitude>49.309837</Latitude> <Longitude>-123.082108</Longitude> <Region>British Columbia</Region> <Locality>North Vancouver</Locality> <Country>Canada</Country>
DISTANCE	1.17 Km

Table 3: Heuristic performance georeferencing experiments configuration.

run	Parameters		
	Geo. Heuristic	StopWords	Dictionary
Ex1	population (pop)	no	no
Ex2	population	yes	yes
Ex3	knowledge (know)	no	no
Ex4	knowledge	yes	yes
Ex5	know+pop	no	no
Ex6	know+pop	yes	yes

Table 4: Results of the heuristic performance georeferencing experiments with the MediaEval 2010 Placing Task test data (5091 videos).

run	#pred.OK	#predict.	Accuracy	Precision
	100km			
Ex1	2185	4681	0.429	0.466
Ex2	2337	4173	0.459	0.560
Ex3	1958	2353	0.384	0.832
Ex4	1919	2215	0.376	0.863
Ex5	2732	4681	0.536	0.583
Ex6	2839	4173	0.557	0.680

ulation heuristics. applied alone or both combined with prioritizing first the geographical knowledge ones. The results (see Table 4) show that the heuristics that apply geographical knowledge without population heuristics obtain the best precision with a 86.36% of correctly predicted videos from the 2215 predicted in the experiment *Ex4* and a 83.21% of correctly predicted videos from the 2353 predicted in the experiment *Ex3*. The difference between *Ex3* and *Ex4* is that the last one uses stopwords and the English dictionary to filter out ambiguous place names. On the other hand the combination of geographical knowledge and population heuristics in experiments *Ex5* and *Ex6* obtained a precision of 58.36% (with 4681 predicted videos) and 68.03% (with 4173 predicted videos) respec-

tively. In order to know the relative performance in precision of each specific heuristic that pertains to the geographical knowledge set of heuristics (described in Section 3.2) we computed the precision of each rules (applied in priority order) in the context of the experiment *Ex4* (see Table 5).

Table 5: Relative performance in precision of each geographical knowledge heuristic data set georeferencing up to 100 Km with the experiment *Ex4*.

Heuristic Feature (Superord.)	Measures		
	Pred.OK	Pred.	Precision
H1_city/spot (state)	1351	1546	0.873
H2_city/spot (country)	515	609	0.845
H3_state (country)	53	60	0.883

5 Language Modelling and Information Retrieval

We used a Language Modelling approach based on statistical Language Models (LM) and Information Retrieval (IR) algorithms to georeference users textual annotations. We used the Terrier⁸ IR software platform to do the experiments with some of their provided IR and LM algorithms. The IR models used for the experiments were the TFIDF and the Okapi BM25 weighting model. The Language Model used was the Hiemstra Language Modelling (HLM) approach (Hiemstra, 2001). The parameters selected were the default used in Terrier (version 3.0). For the Hiemstra LM the default lambda value of 0.15 was used (Hiemstra, 2001).

The models were trained with a metadata corpus of Flickr photos provided in the MediaEval 2010 for development purposes. The corpus consists of 3,185,258 Flickr pho-

⁸Terrier. <http://terrier.org>

tos uniformly sampled from all parts of the world. The photos are georeferenced with geotags with 16 zoom accuracy levels. The accuracy shows the zoom level used by the user when placing the photo on the map ((e.g., 6 - region level, 12 - city level, 16 - street level). The metadata of the corpus is represented by the following information: *UserID*, *PhotoID*, *HTMLinkToPhoto*, *GeoData* (includes longitude, latitude, and zoom accuracy level), *tags*, *date taken*, and *date uploaded*. From the metadata corpus of photos we filtered out some data: 1) if a user has several photos metadata with the same tagset we kept only one photo metadata of them, 2) we filtered out metadata without existing tags. After this filtering steps we got a set of 1,723,090 metadata entries for each photo. Then, from the filtered corpus we selected four subsets depending on the values of the zoom level accuracy: 1) level 16 (715,318 photos), 2) levels from 14 to 16 (1,140,031 photos), 3) levels from 12 to 16 (1,570,771 photos), 4) levels from 6 to 16 (1,723,090 photos). Moreover, for each unique coordinates pair in the corpora we joined all the tagsets associated to the same coordinates resulting of: 1) level 16 (511,222 coordinates), 2) levels from 14 to 16 (756,916 coordinates), 3) levels from 12 to 16 (965,904 coordinates), 4) levels from 6 to 16 (1,026,993 coordinates). The indexing of the metadata subsets were done with the coordinates as a document number and their associated tagsets the document text. We indexed with filtering using the multilingual stopwords list and without stemming. The retrieval experiments have been done with a subset of the metadata fields of the videos consisting in Keywords (tags), Title and Description as queries to the IR system. The metadata fields Title and Description were lowercased for the query. Given a query to the IR system, the first one top-ranked retrieved coordinate pair is selected as the prediction for the video. The experiments shown in Table 6 show that BM25 achieves the best results in accuracies from 10 to 100 Km and the Hiemtra LM achieves the best results in accuracies georeferencing up to 1 and 5 km.

6 Re-Ranking and GeoFusion

A Re-Ranking (RR) process is applied after the IR or LM retrieval. For each topic (set of metadata fields, keywords, description and

Table 6: Results of the georeferencing experiments with Language Modelling.

Model	Accuracy				
	1km	5km	10km	50km	100km
(A) annotation accuracy=16					
BM25	0.423	0.505	0.539	0.595	0.609
TFIDF	0.422	0.502	0.536	0.591	0.605
HLM	0.430	0.505	0.535	0.598	0.613
(B) annotation accuracy=14-16					
BM25	0.423	0.506	0.544	0.599	0.612
TFIDF	0.422	0.504	0.541	0.593	0.606
HLM	0.436	0.512	0.547	0.607	0.621
(C) annotation accuracy=12-16					
BM25	0.420	0.504	0.551	0.606	0.621
TFIDF	0.420	0.504	0.549	0.602	0.617
HLM	0.435	0.514	0.551	0.604	0.620
(D) annotation accuracy=6-16					
BM25	0.414	0.501	0.552	0.606	0.624
TFIDF	0.413	0.501	0.550	0.602	0.620
HLM	0.428	0.510	0.549	0.604	0.622

title associated to a photo of the MediaEval 2010 test set) their first 1000 retrieved coordinates pairs from the IR software are used. From them we selected the subset of coordinates pairs with a weight equal or greater than the 66.66% of the weight of the coordinates pair ranked in first position. Then for each geographical coordinates pair of the subset we sum its associated weight (provided by the IR or LM algorithm) and the weight of their Nearest Neighbours at different distances of 1, 5, 10, 50, and 100 km. Then we select the one with the maximum weight sum.

The GeoFusion approach is applied by combining the results of the Geographical Knowledge approach and the Language Modelling approach with Re-Ranking (see the results of this approach in Table 7). The results are combined in the following way: from the set of Geographical based experiments we selected the experiment with best precision (*Ex4*). From the Geographical Knowledge experiment with highest precision the system selects the predicted coordinates, and the ones that are not predicted due to the geographical rules do not match are selected from the Language Modelling (LM) approaches with Re-Ranking. This means that from the *Ex4* were selected 2215 predictions and the rest (2876 predictions) were selected from the LM with RR approaches. The results of the Re-Ranking and the GeoFusion approaches (see Table 7) show that

both approaches outperform the Geographical and the Language Model approaches and the baselines. The baselines presented in Table 7 are three: 1) the best results obtained at the Placing Task of MediaEval 2010 with the test set (Laere, Schockaert, and Dhoedt, 2010a), 2) the experiment with BM25 trained with a corpus with accuracies from 6 to 16 levels, and 3) the Hiemstra LM trained with accuracies from 14 to 16 levels. These last two baselines were the ones that obtained the best results in accuracies compared to the other Language Models and corpus training models (see Table 6).

Table 7: Results of the georeferencing experiments with Geographical Knowledge and Language Modelling with and without Re-Ranking

Model	Accuracy				
	1km	5km	10km	50km	100km
Baselines					
Laere2010	0.432	0.542	0.587	0.650	0.672
BM25 (D)	0.414	0.501	0.552	0.606	0.624
HLM (B)	0.436	0.512	0.547	0.607	0.621
cluster distance=1km					
BM25_RR	0.433	0.513	0.550	0.605	0.623
+Geo	0.259	0.454	0.524	0.630	0.655
HLM_RR	0.453	0.533	0.569	0.633	0.649
+Geo	0.272	0.467	0.539	0.649	0.673
cluster distance=5km					
BM25_RR	0.369	0.526	0.563	0.621	0.637
+Geo	0.242	0.463	0.533	0.638	0.664
HLM_RR	0.403	0.543	0.573	0.646	0.659
+Geo	0.254	0.476	0.547	0.659	0.682
cluster distance = 10km					
BM25_RR	0.368	0.505	0.577	0.625	0.639
+Geo	0.242	0.456	0.538	0.640	0.666
HLM_RR	0.403	0.527	0.589	0.648	0.661
+Geo	0.256	0.470	0.552	0.661	0.684
cluster distance = 50 Km					
BM25_RR	0.349	0.468	0.512	0.634	0.647
+Geo	0.230	0.437	0.514	0.643	0.668
HLM_RR	0.383	0.492	0.542	0.648	0.659
+Geo	0.243	0.455	0.534	0.659	0.683
cluster distance = 100 km					
BM25_RR	0.350	0.463	0.500	0.595	0.648
+Geo	0.230	0.439	0.511	0.631	0.670
HLM_RR	0.383	0.490	0.535	0.618	0.660
+Geo	0.243	0.453	0.529	0.646	0.685

7 Conclusions

We presented generic georeferencing approaches based on Geographical Knowledge Bases, Linguistic Knowledge, and Language Modelling to deal with users textual annota-

tions and tags. We also analyzed and discussed some of the many problems to face georeferencing textual annotations and tags and its possible resolution. The main contribution of this paper is to show that Geographical Knowledge and Language Modelling can be wisely combined for achieving and improving state-of-the art results in georeferencing (both in accuracy and precision). Our experiments show that stop-words lists and controlled dictionaries can help the disambiguation of placing names and the focus detection. The experiments also show that geographical knowledge heuristics can achieve a high precision in georeferencing: up to a 86.36%. This fact is very interesting for establishing high confidence rules that could allow a high precision georeferencing detection in textual annotations and tags. The strategy that combines geographical knowledge and population heuristics for geographical focus detection achieves the best results in the experiments with the Geographical approach with the MediaEval 2010 Placing Task data set. The Language Modelling approaches outperformed the Geographical one, but the fusion of both is achieving the best results. The best approach georeferencing up to 1, 5 and 10 km is achieved with the Language Modelling Re-ranking approach with the Hiemstra LM. The best results in accuracy up to 50 and 100Km are achieved with the fourth strategy: a fusion of Language Modelling Re-ranking with Geographical Knowledge approaches. These strategies outperformed some the best results in accuracy reported by the state-of-the art systems participating at MediaEval 2010 Placing task. Our best results achieved are 68.53% of accuracy georeferencing up to a distance of 100 Km. The approaches of (Laere, Schockaert, and Dhoedt, 2010a) and (Kelm, Schmiedeke, and Sikora, 2010) obtained a 67,23% and 60,46% of accuracy with the same test set at the MediaEval 2010 placing task.

Further work will include: 1) the use of Geographical Feature Types Thesaurus and Geographical Names language-dependent patterns, 2) new experiments with Re-Ranking algorithms based on distances between coordinates pairs, 3) the study of the social information present in the metadata and how could be used for improving georeferencing.

References

- Adrian Popescu, Pierre-Alain Moëllic, Ioannis Kanellos. 2008. Themeexplorer: Finding and Browsing Geo-Referenced Images. In *Proceedings of the International Workshop Content-Based Multimedia Indexing*, pages 576–583, London, United Kingdom.
- Crandall, David J., Lars Backstrom, Daniel Huttenlocher, and Jon Kleinberg. 2009. Mapping the World’s Photos. In *Proceedings of the 18th international conference on World wide web, WWW ’09*, pages 761–770, New York, NY, USA. ACM.
- Ferrés, Daniel and Horacio Rodríguez. 2010. TALP at MediaEval 2010 Placing Task: Geographical Focus Detection of Flickr Textual Annotations. In *Working Notes of the MediaEval 2010 Workshop, October 24, 2010*, Pisa, Italy.
- Geoffrey Andogah, Gosse Bouma, John Nerbbonne, Elwin Koster. 2008. Geographical Scope Resolution. In *Methodologies and Resources for Processing Spatial Language, Workshop at the 6th Conference on Language Resources and Evaluation (LREC)*, pages 4–10, Marrakech, Morocco.
- Hauptmann, Alexander G., Er G. Hauptmann, and Andreas M. Olligschlaeger. 1999. Using Location Information From Speech Recognition Of Television News Broadcasts. In *Proceedings of the ESCA ETRW Workshop on Accessing Information in Spoken Audio*, pages 102–106, Cambridge, England. University of Cambridge.
- Hays, James and Alexei A. Efros. 2008. im2gps: Estimating Geographic Information from a Single Image. In *Proceedings of the IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*.
- Hiemstra, Djoerd. 2001. *Using Language Models for Information Retrieval*. Ph.D. thesis, Enschede.
- Kelm, Pascal, Sebastian Schmiedeke, and Thomas Sikora. 2010. Video2GPS: Geotagging Using Collaborative Systems, Textual and Visual Features. In *Working Notes of the MediaEval 2010 Workshop, October 24, 2010*, Pisa, Italy.
- Keßler, Carsten, Patrick Maué, Jan Torben Heuer, and Thomas Bartoschek. 2009. Bottom-Up Gazetteers: Learning from the Implicit Semantics of Geotags. In *GeoS*, pages 83–102.
- Laere, Olivier Van, Steven Schockaert, and Bart Dhoedt. 2010a. Ghent University at the 2010 Placing Task. In *Working Notes of the MediaEval 2010 Workshop, October 24, 2010*, Pisa, Italy.
- Laere, Olivier Van, Steven Schockaert, and Bart Dhoedt. 2010b. Towards Automated Georeferencing of Flickr Photos. In *Proceedings of the GIR’10 Workshop*, Zurich, Switzerland.
- Larson, M., M. Soleymani, P. Serdyukov, S. Rudinac, C. Wartena, V. Murdock, G. Friedland, R. Ordelman, and G. J. F Jones. 2011. Automatic Tagging and Geotagging in Video Collections and Communities. In *Proceedings of the ACM International Conference on Multimedia Retrieval (ICMR 2011)*.
- Leidner, J. L. 2007. *Toponym Resolution: a Comparison and Taxonomy of Heuristics and Methods*. PhD Thesis, University of Edinburgh.
- Perea-Ortega, José M., Miguel Á. García-Cumbreras, L. Alfonso Ure na López, and Manuel García-Vega. 2010. SINAI at Placing Task of MediaEval 2010. In *Working Notes of the MediaEval 2010 Workshop, October 24, 2010*, Pisa, Italy.
- Popescu, Adrian and Ioannis Kanellos. 2009. Creating Visual Summaries for Geographic Regions. In *ECIR Workshop on Information Retrieval over Social Networks*.
- Serdyukov, Pavel, Vanessa Murdock, and Roelof van Zwol. 2009. Placing Flickr Photos on a Map. In James Allan, Javed A. Aslam, Mark Sanderson, ChengXiang Zhai, and Justin Zobel, editors, *SIGIR*, pages 484–491.
- Stokes, Nicola, Yi Li, Alistair Moffat, and Jiawen Rong. 2008. An Empirical Study of the Effects of NLP Components on Geographic IR performance. *International Journal of Geographical Information Science*, 22(3):247–264.